Open Data of the LHC ATLAS experiment

Jiří Chudoba

FZU Fyzikální ústav Akademie věd České republiky

20th Conference of Czech and Slovak Physicists, September 2020


SUISSE
FRANCE
LHCb
CERN
Prévessin
ATLAS
CERN Meyrin
ALICE
CMS
LHC 27 km

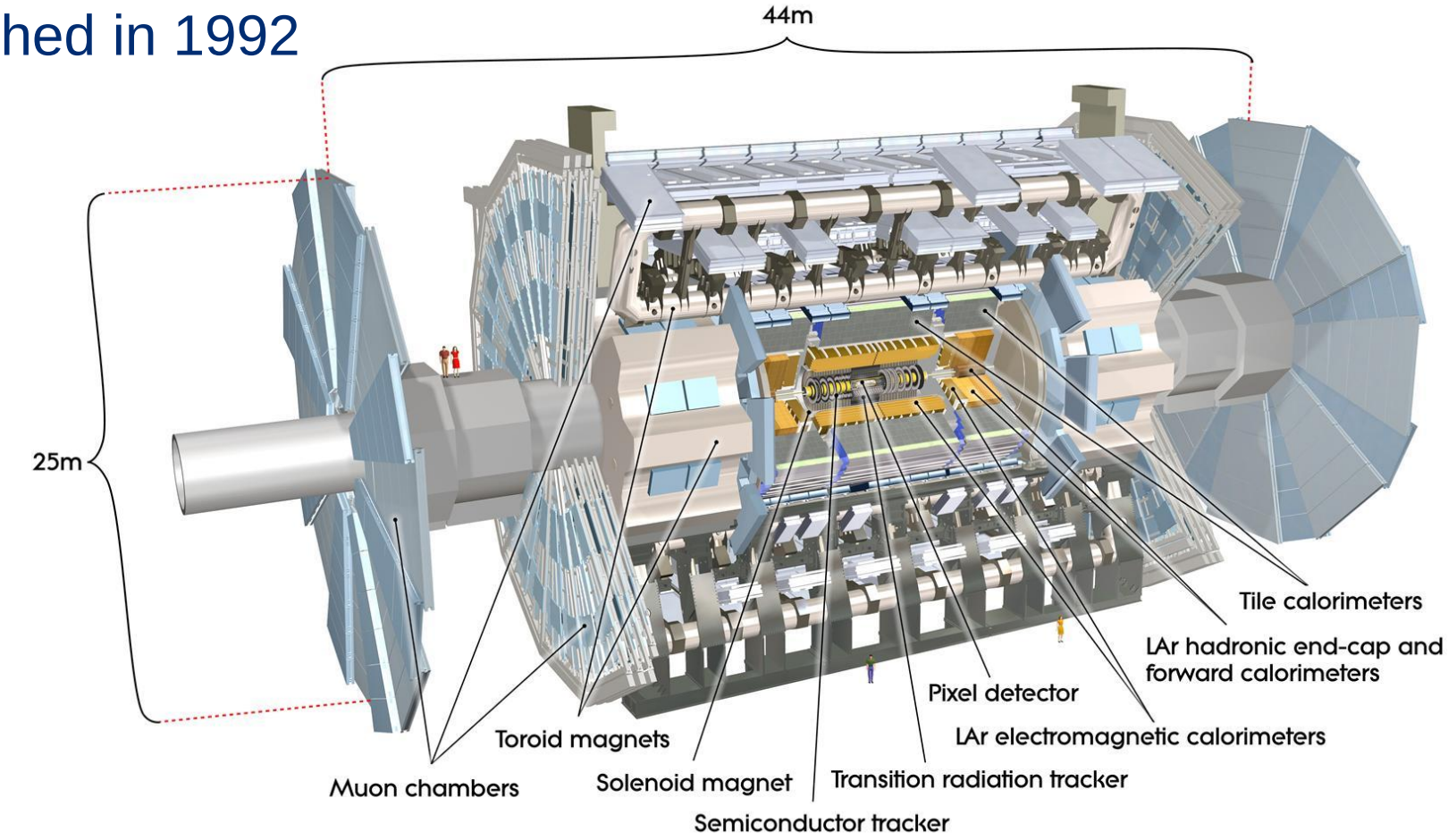
ATLAS experiment

Collaboration established in 1992

38 countries

181 institutions

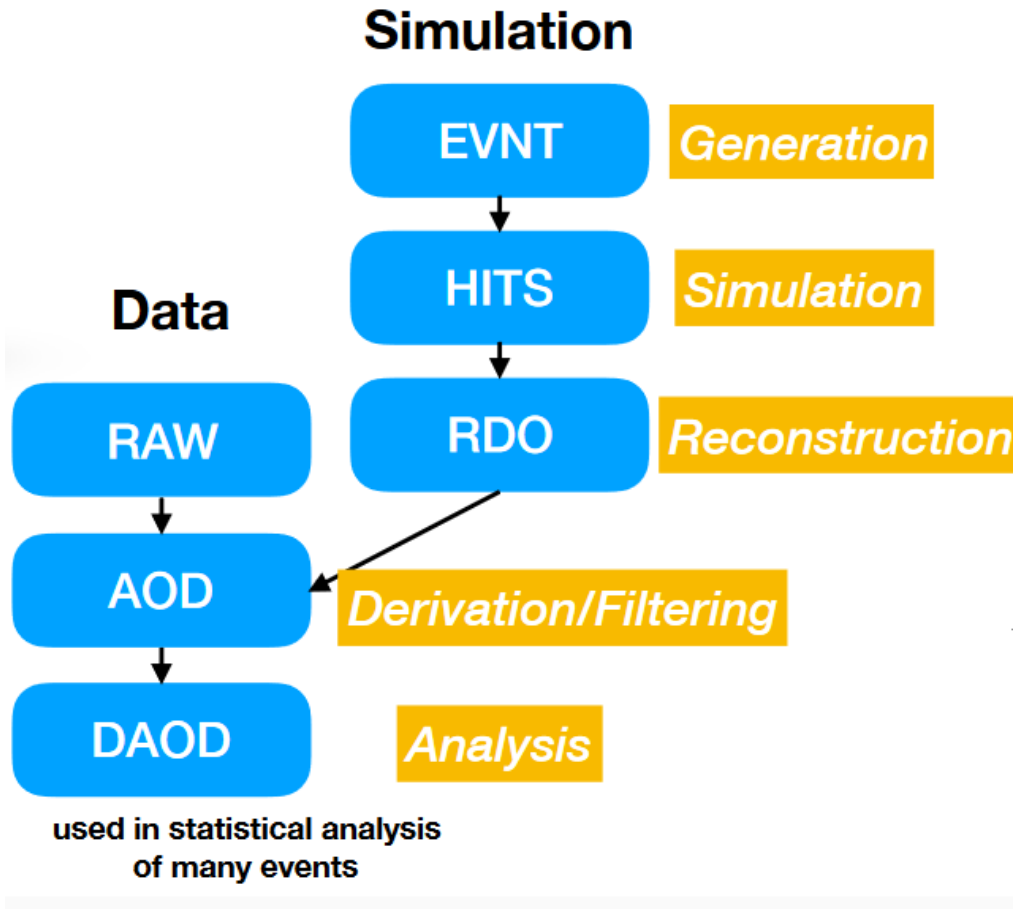
3000 authors



ATLAS Data

- ATLAS raw: 1 MB/event
- Raw total 57 PB (Run 1 and 2)
- 211 PB (disk) + 275 PB (tape)

.5 EB = 500 000 000 GB

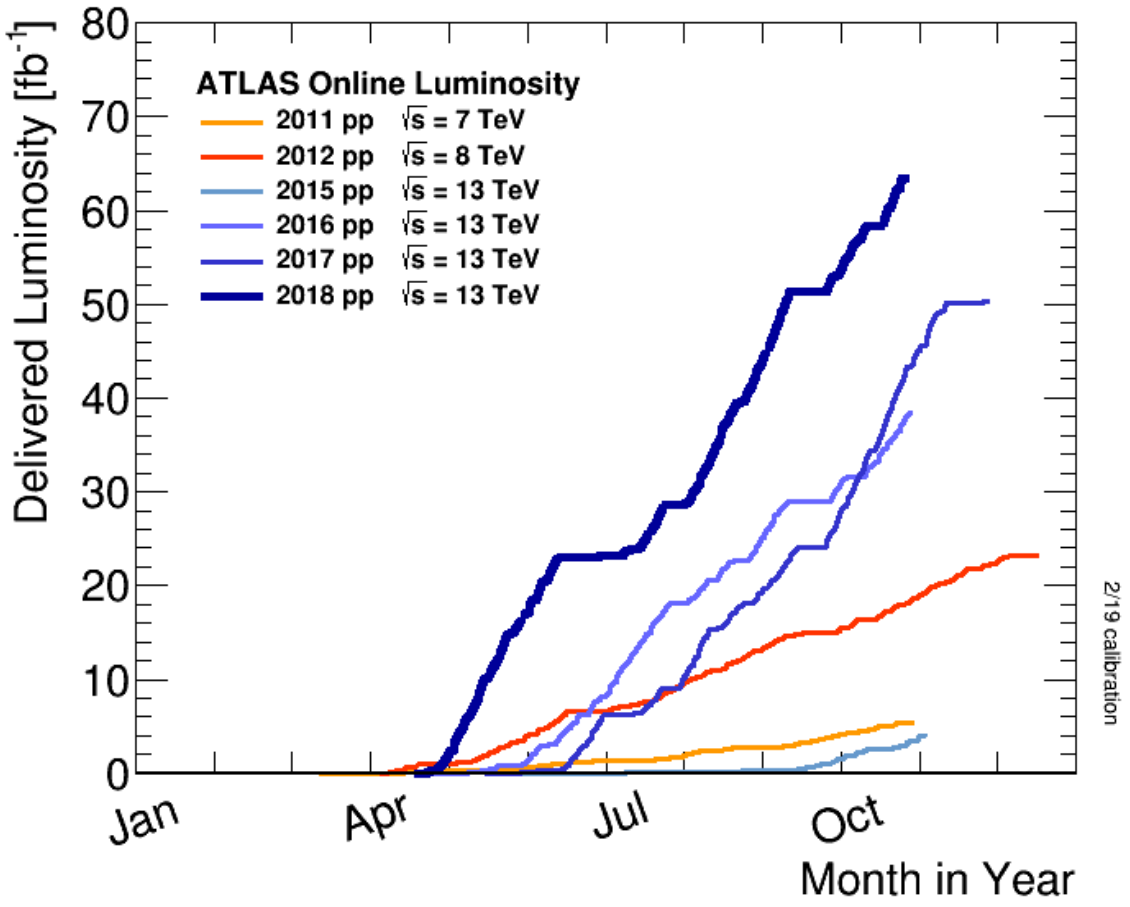


Open Data

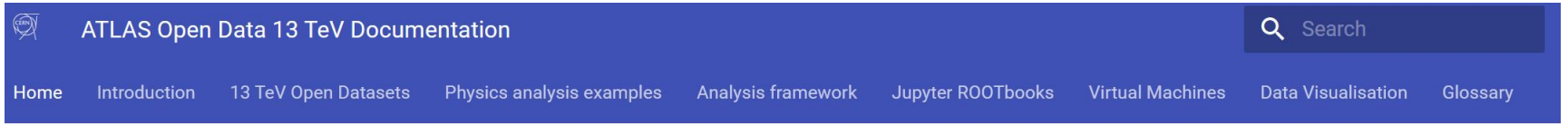
- Aims: **Education and Outreach**, Science
  - Only final data formats after data processing

ATLAS Open Data Datasets

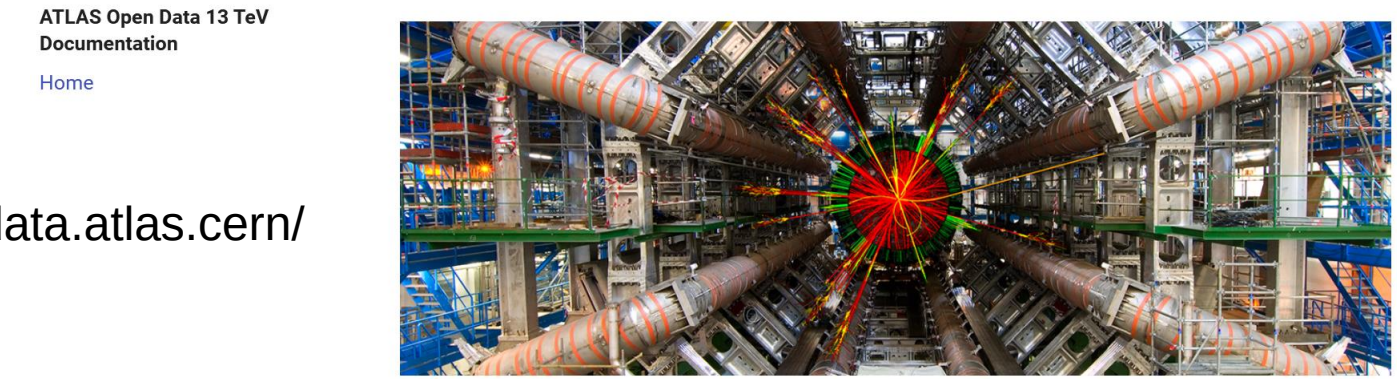
- 8 TeV data from 2012
  - 1 fb^{-1}
  - MC: 44 samples
- 13 TeV data from 2016
  - 10 fb^{-1}, 28 GB
  - 270 million collision events
  - MC: 120 samples, 88 GB
  - **116 GB total**



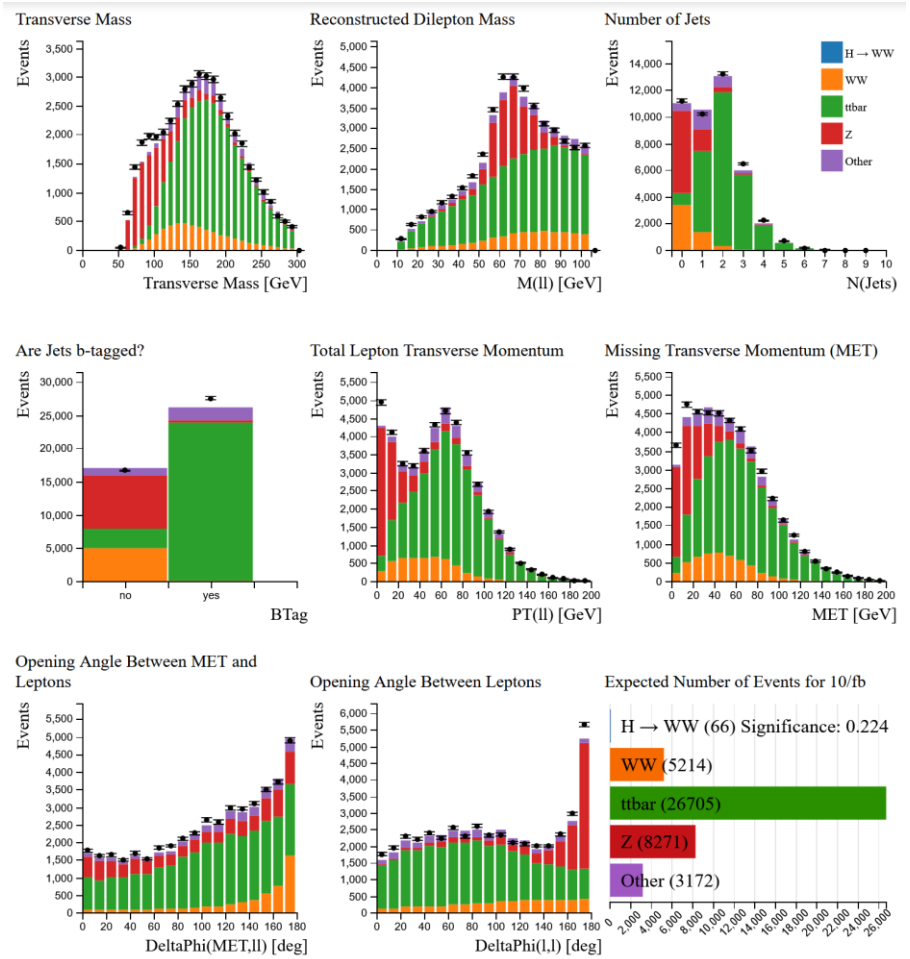
ATLAS Open Data Portal

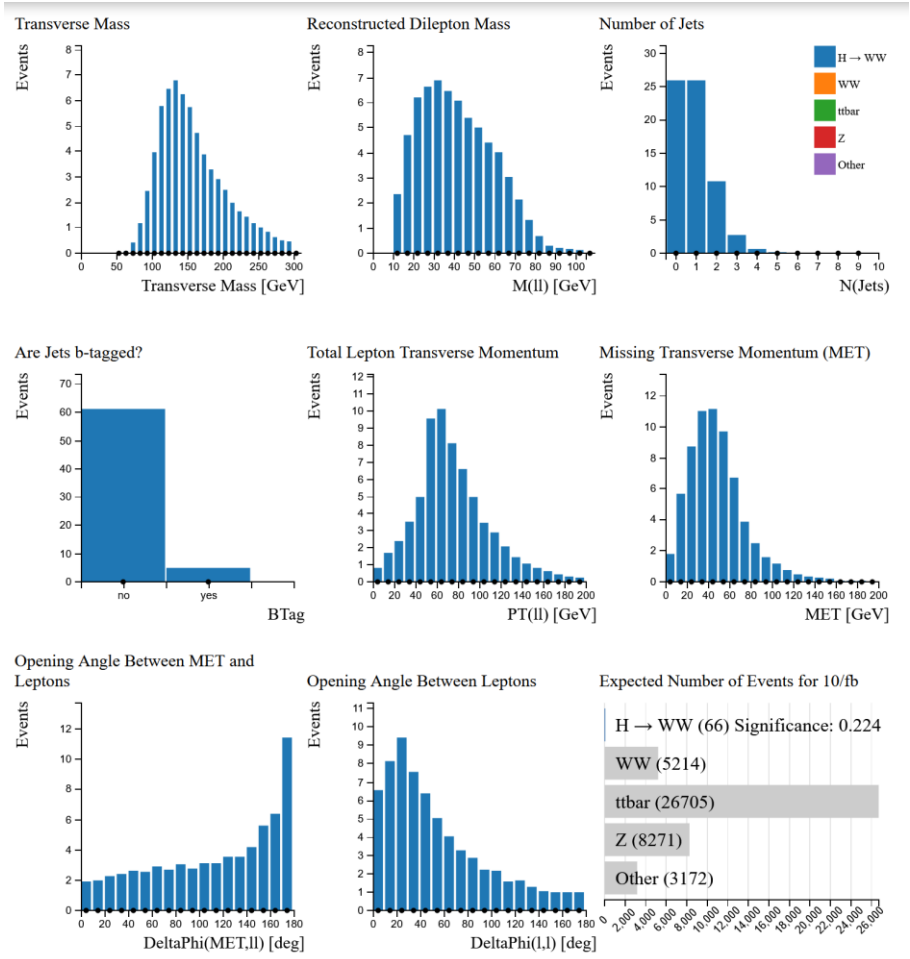


http://opendata.atlas.cern/



Tools: Interactive Histograms I

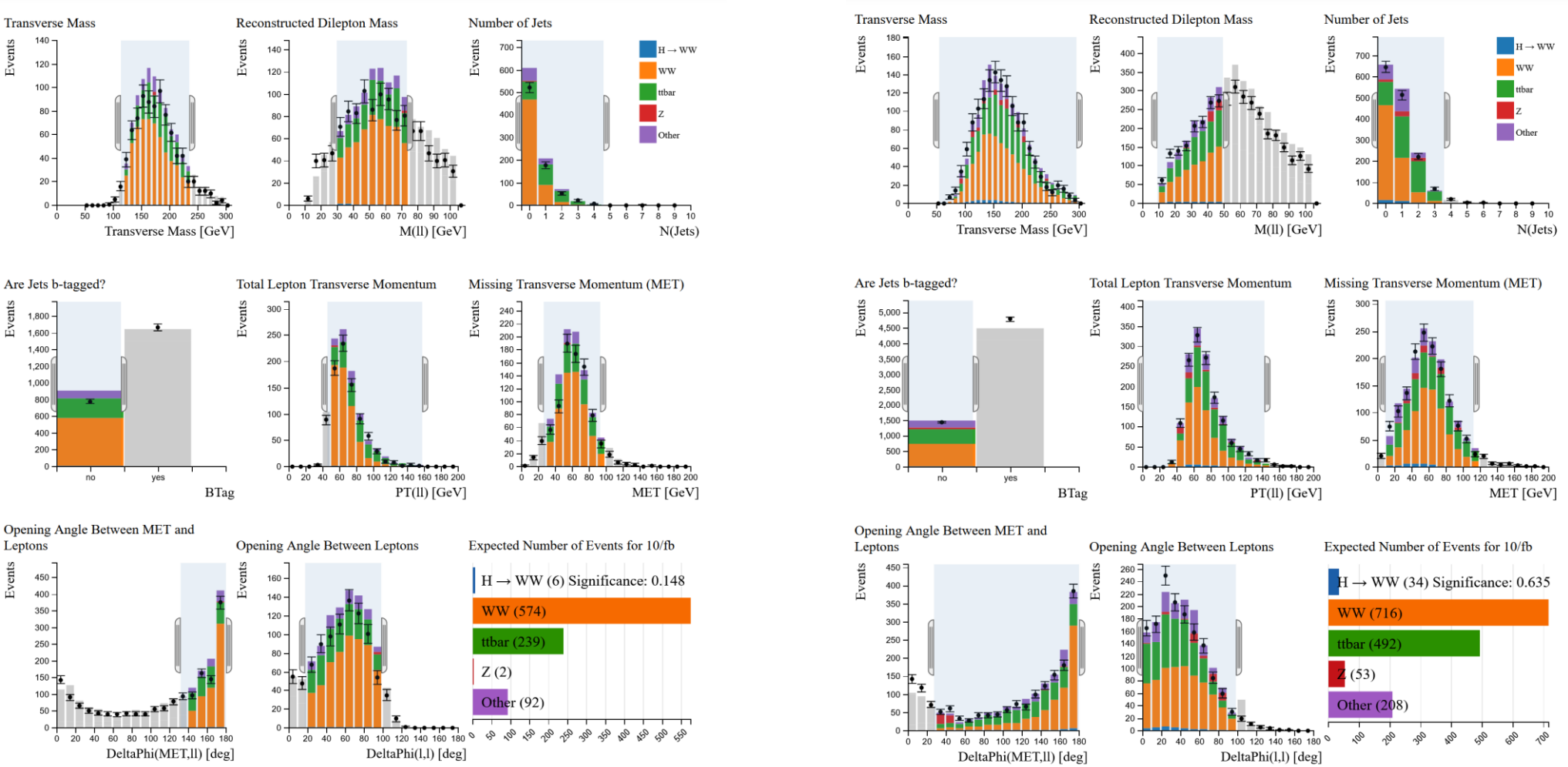




FZU Fyzikální ústav Akademie věd České republiky

Jiri.Chudoba@cern.ch

Tools: Interactive Histograms II

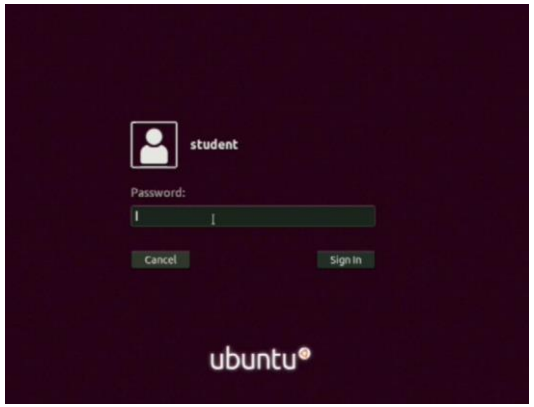


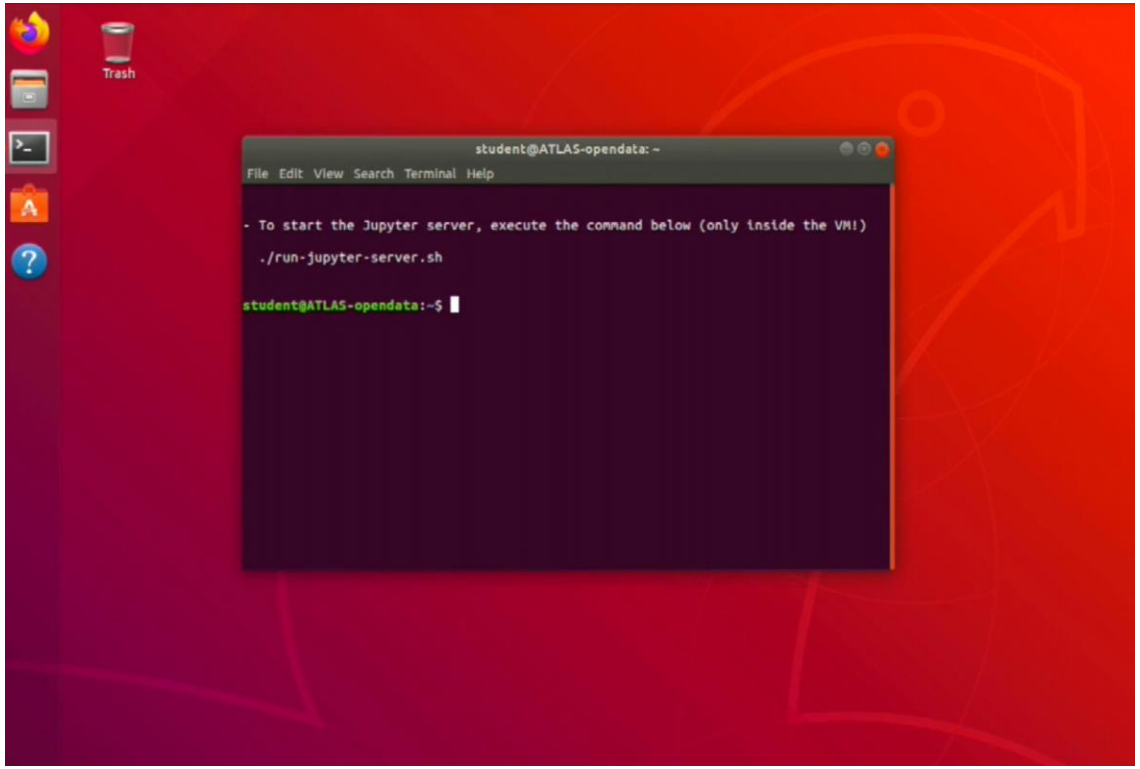
FZU Fyzikální ústav Akademie věd České republiky

Jiri.Chudoba@cern.ch

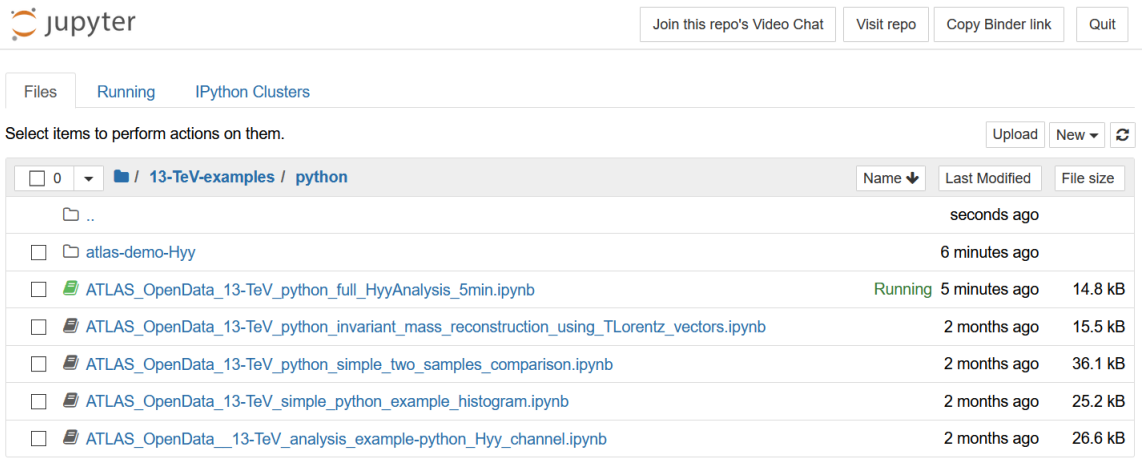
Tools: Virtual Machines

- Windows vs MacOS vs Linux: VirtualBox images for download
- Several sizes (with more or less data)
- Video tutorials





Jupyter Notebooks



Searching for the Higgs boson in the H→γγ channel

C++ notebook example

Introduction Let's take a current ATLAS Open Data sample and create a histogram:

In order to activate the interactive visualisation of the histogram that is later created we can use the JSROOT magic:

In [1]:
```
%jsroot on
```

We need to include some standard C++ and ROOT libraries

In [2]:
```
// Creates a TChain to be used by the Analysis.C class
#include <TChain.h>
#include <vector>
#include <TFile.h>
#include <iostream>
#include <string>
#include <stdio.h>
```

Because we would like to use more than one ROOT input file, the best option is to use a TChain object. This allows to "chain" several samples into a single structure that we can later loop over

In [3]:
```
TString path = "https://atlas-opendata.web.cern.ch/atlas-opendata/samples/2020/GamGam/"
```

```
(TString &) "https://atlas-opendata.web.cern.ch/atlas-opendata/samples/2020/GamGam/"[70]
```

In [4]:
```
TChain* fChain = new TChain("mini");

fChain->AddFile(path+"Data/data_A.GamGam.root");
fChain->AddFile(path+"Data/data_B.GamGam.root");
fChain->AddFile(path+"Data/data_C.GamGam.root");
fChain->AddFile(path+"Data/data_D.GamGam.root");
```

Binder

Binder allows you to create custom computing environments that can be shared and used by many remote users. It is powered by BinderHub, which is an open-source tool that deploys the Binder service in the cloud. One-such deployment lives here, at mybinder.org, and is free to use.

https://hub.gke.mybinder.org/user/atlas-outreach--ection-ope

jupyter ATLAS_OpenData_13-TeV_python_full_HyyAnalysis_5min (autosaved) Join this repo's Video Chat

File Edit View Insert Cell Kernel Widgets Help Notebook saved Trusted Python 3

Run Markdown Download **Memory:** 1397 / 2048 MB

GitHub Binder

Clone the analysis code **(do this only once!)**

```
In [10]: myCmd = os.popen('git clone https://github.com/atlas-outreach-data-tools/atlas-outreach-cpp-fram
         print(myCmd)
```

```
Cloning into 'atlas-outreach-cpp-framework-13tev'...
```

Get into the folder which contains the analysis code

```
In [11]: os.chdir("atlas-outreach-cpp-framework-13tev/")
```

```
In [12]: myCmd = os.popen('ls -lhrt').read()
         print(myCmd)
```

```
total 28K
drwxr-xr-x 14 jovyan root 4.0K Sep  9 07:55 Analysis
-rwxr-xr-x  1 jovyan root  370 Sep  9 07:55 welcome_web.sh
-rwxr-xr-x  1 jovyan root 1.7K Sep  9 07:55 welcome.sh
-rw-r--r--  1 jovyan root  11K Sep  9 07:55 README.md
drwxr-xr-x  4 jovyan root 4.0K Sep  9 07:55 Plotting
```

https://hub.gke.mybinder.org/user/atlas-outreach--ection-ope

jupyter ATLAS_OpenData_13-TeV_python_full_HyyAnalysis_5min (autosaved) Join this repo's Video Chat

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Markdown Download **Memory:** 1352 / 2048 MB

GitHub Binder

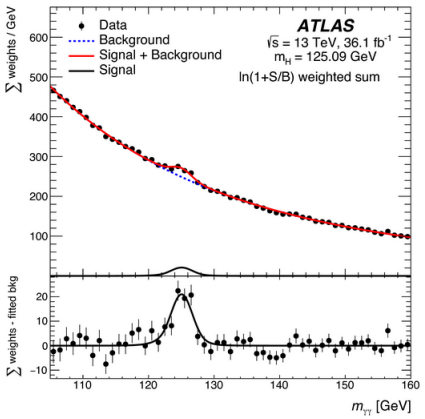
The two photons invariant mass

mass

```
In [28]: end = time.time()
         duration = end-start
         print("Finished in {} min {} s".format(int(duration//60),int(duration%60))) # Python3
```

```
Finished in 10 min 51 s
```

Finally, you can compare it to the ATLAS plot from a published paper that uses **3 times more data!**



FZU Fyzikální ústav Akademie věd České republiky

Jiri.Chudoba@cern.ch

Analysis Examples

Hyy Analysis

Based on:

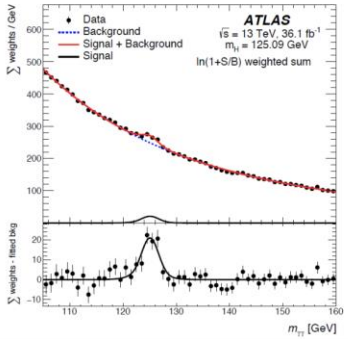
PRD 98 (2018) 052005 11th January 2019

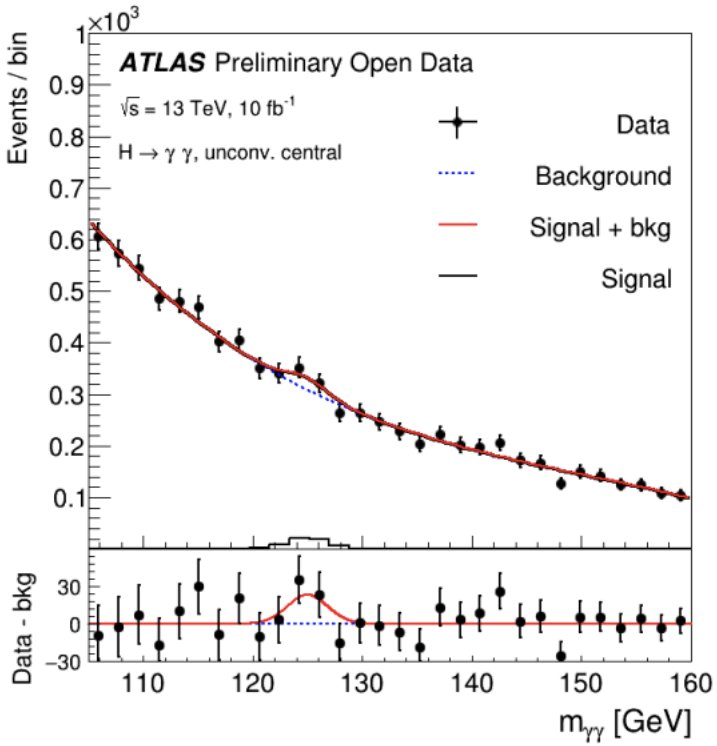
Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV with the ATLAS detector

Simplified selection:

- 2 photons with pT > 35, 25 GeV
- E_T / m(γγ) > 0.35(0.25)
- 105 GeV < m(γγ) < 160 GeV

Figures to reproduce:





Kate Shaw, ICHEP2020

Documentation

Example: Find the Higgs

Follow the steps of a real ATLAS analysis using Histogram Analyser

$H \to W^+W^- \to \ell^+\ell^-\nu\bar{\nu}$

(ℓ = lepton = electron or muon)

We are looking for a Higgs boson which decays into two W bosons which subsequently decay into leptons and neutrinos.

The major background contributions to the search in this decay mode are top (top quark pair and W+top quark), WW and Z+jets events.

Use the cursor to implement the following cuts, one by one.
As you apply the cuts, you should see the distributions changing.

Glossary Search

Invariant mass

Particle physicists use the word "mass" to refer to the quantity (sometimes called "rest mass") which is proportional to the inertia of the particle when it is at rest. When a particle decays and hence no longer exists, its mass before the decay can be calculated from the energies and momenta of the decay products. The inferred value of the mass is independent of the reference frame in which the energies and momenta are measured, so that that the mass called "invariant".

Isolation

An object isolation is an important feature, because it indicates where the object (lepton, jet,..) stems from. As an example, an isolated electron is likely to originate from a Z-boson, W-boson or tau-lepton. In contrast, a non-isolated electron is likely to originate from a b-hadron or c-hadron decay.

Jets

Jets are the dominant final state objects of high-energy proton-proton interactions at the LHC. They are key ingredients for many physics measurements and for searches for new phenomena. Jets are observed as groups of topologically-related energy deposits in the ATLAS calorimeters, most of which are associated with tracks of charged particles as measured in the inner detector. They are reconstructed and calibrated using a combination of methods based on simulation and data-driven techniques.

Jet cleaning criteria have been developed in order to identify fake jets which arise due to noise or to out-of-time energy depositions. Jets failing these criteria are flagged as either "bad", likely to be fake, or "ugly", likely to be mismeasured due to falling into less well instrumented regions.

http://opendata.atlas.cern/release/2020/documentation/atlas/GLOSSARY.html

Program Open Science

- Organized by SSČ AV ČR for secondary school students
- Various topics
- 1 to 3 students per topic
- 8 hours/month for 1 year
- Concluded by Conference

Open Science Experience

- Many new terms and tools
  - Virtual Machines, Cloud Computing
  - Linux
  - Bash, Python, ROOT, C++
  - Jupyter Notebooks
  - Git and Github
  - GitBook
  - Virtual Reality

Conclusion

ATLAS Open Data provide excellent tool for education

http://opendata.atlas.cern

Thank you for your attention!